A ROBUST 2400bit/s MBE-LPC SPEECH CODER INCORPORATING JOINT SOURCE AND CHANNEL CODING

D. Rowe and P. Secker

Department of Electrical and Computer Engineering
The University of Wollongong, P.O. Box 1144, Wollongong, NSW 2500, Australia

ABSTRACT

This paper presents a robust speech coder based on the Multi-Band Excitation (MBE) model. Unlike many existing MBE coders, the spectral envelope information is represented using Linear Prediction Coefficients (LPCs), which are quantised and encoded using a joint source and channel trellis coding scheme to obtain remarkable robustness to random channel errors without increase in bit rate (i.e. no redundant forward error correction (FEC) required). The coder produces communications quality speech at 2400 bit/s and can maintain acceptable speech quality at bit error rates (BERs) up to 5%. The complexity is low enough for implementation on commercial DSP devices.

1.0 INTRODUCTION

The Multi-Band Excitation model is a compact speech coding model suitable for the medium to low bit rate (8000 bit/s to 1500 bit/s) transmission of communications quality speech. It is thus well suited to bandwidth and power critical services such as mobile satellite communications [1]. These channels are characterized by random and burst errors, thus the requirement for robust low bit rate speech coding technology.

The MBE model [2] represents speech as the multiplication of a spectral envelope and an excitation spectrum. The excitation spectrum contains both voiced and unvoiced regions. The voiced/unvoiced decisions are made over each harmonic of the fundamental frequency. The MBE model components are spectral envelope, fundamental frequency, and a voiced/unvoiced decision for each harmonic of the fundamental frequency. The spectral amplitude information is represented using a variable number of discrete spectral amplitude samples, one sample for each band. Note that the number of bands, and hence the number of spectral amplitude samples and voiced/unvoiced decisions is dependant on the fundamental frequency.

Section 2 of this paper describes the transformation used to convert the spectral amplitude samples to the LPC model. Section 3 presents the techniques used to quantise and protect the LPC information using joint source and channel coding techniques. The fully quantised coder is presented and subjectively evaluated in random error channels in section 4.

2.0 THE MULTIBAND EXCITATION LINEAR PREDICTIVE SPEECH CODER

Traditional MBE speech coders [1][2][3] have relied on adaptive bit allocation techniques to represent the spectral amplitude information, the bit allocation for each frame being dependant on the fundamental frequency. One undesirable property of these approaches is the sensitivity of the coder to channel bit errors in the fundamental frequency. The MBE-LPC speech coder [4][5][6] transforms the MBE model spectral amplitude samples to a fixed number of Linear Prediction Coefficients (LPCs) which are then quantised using LSPs.

This scheme has several advantages including fixed bit allocation and higher robustness to channel bit errors due to the removal of the fundamental frequency dependant bit allocation. Algorithmic complexity and storage requirements are also substantially reduced when compared to other MBE spectral amplitude quantisation techniques, simplifying real time implementation.

2.1 Spectral Amplitude Transformation

The MBE model parameters are derived from the input speech signal using a combination of time and frequency domain methods based on the Inmarsat-M Improved Multi-Band Excitation voice coding specification [1]. The spectral amplitude information is in the form of discrete amplitude samples $\{A_m\}$ equally spaced along the frequency axis at intervals of ω_o, the fundamental frequency There is one spectral amplitude sample for each band. These samples represent the smoothed spectrum (formant structure) of the speech. This MBE model spectral amplitude information is thus described by the spectral amplitude samples and ω_o. The model parameters are updated once every 30ms.

The spectral amplitude samples A_m can be considered to be discrete Fourier transform magnitude samples containing the spectral envelope information for the current frame of speech.

The LPC model is a popular method of modelling speech spectra using a small number of parameters [7][8]. The LPC synthesis model commonly used for speech consists of an excitation source $E(z)$, and a spectral shaping filter $H(z) = 1/A(z)$ where:

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{1}$$

where $\{a_k\}$ are the linear prediction coefficients describing the filter, and p is the linear prediction order. A linear prediction order of 10 has been chosen for this application.

The procedure used to transform a variable number of spectral amplitude samples A_m to a fixed number of LPC coefficients $\{a_k\}$ is to first determine the first p+1 autocorrelation coefficients $\{R_i\}$ of the spectrum described by $\{A_m\}$. This can be achieved using a discrete Fourier transform (DFT) and the Wiener-Khinchin theorem [7] as:

$$R_i = \frac{1}{M} \sum_{m=0}^{M-1} A_m^2 \cos(i\omega_o) \tag{2}$$

where ω_o is the fundamental frequency, and M is the total number of spectral amplitude samples. The LPC coefficients can then be determined using the Levinson-Durbin recursion based on the autocorrelation coefficients, R_i. One other parameter is needed to describe the spectral envelope using the LPC model, this is an energy (gain) term G that conveys the LPC model error energy.

To transform the LPC model representation $\{a_k\}$ of the spectral envelope back to the spectral amplitude sample representation the LPC synthesis filter spectrum is sampled at the appropriate frequency points (3) where $\{A_m'\}$ are the reconstructed spectral amplitude samples.

$$A_m' = \frac{\sqrt{G}}{|A(e^{jw_o m})|} \tag{3}$$

3.0 JOINT SOURCE AND CHANNEL CODING OF SPECTRAL INFORMATION

In a severely bandwidth limited environment, the addition of redundant channel protection codes (FEC) is often made at the expense of source coding information. Thus a tradeoff exists between source code and channel code bit allocation. Within the mobile communication environment, this problem is compounded by the fact that the channel characteristics (specifically the BER) are non-stationary. In an attempt to circumvent the problem of how to apportion bits to source and channel codes and to reduce the system complexity, the combining of source and channel coding into one operation has recently been studied for two powerful source coders; namely the Vector Quantizer [9] and Trellis Coder [10]. By incorporating the expected channel transition probability and by reassigning or modifying the codewords, improved immunity to channel errors is obtained without the inclusion of explicit channel coding. Thus there is no increase in bandwidth due to redundant FEC codes and in many cases the overall system complexity is less.

In [11],[12] the joint source and channel trellis encoder is extended to encode Line Spectrum Pair (LSP) parameters [13]. This representation of LPC coefficients exhibits many favorable deterministic and statistical properties [14] and is commonly used within parametric speech coders (eg. CELP). The coder operates on the p-th order LSP vector. By concatenating p trellises and mapping each component of this vector to one branch of the corresponding trellis a low bit rate coding scheme is

obtained. The Viterbi algorithm is used to search the trellis structure for the path sequence that minimizes the squared error cost function

$$d(\omega_i,\hat{\omega}_{i_j}) = (\omega_i - \hat{\omega}_{i_j})^2 \quad (4)$$

between original i-th LSP ω_i and i-th trellis branch values LSP $\hat{\omega}_{i_j}$.

Dunham and Gray [15] suggest that the cost function be modified to include the distortion due to the channel. Thus (4) is modified to

$$d(\omega_i,\hat{\omega}_{i_j}) = \sum_n (\omega_i - \hat{\omega}_{i_n})^2 \Pr(\hat{\omega}_{i_n}|\hat{\omega}_{i_j}) \quad (5)$$

where $\Pr(\hat{\omega}_{i_n}|\hat{\omega}_{i_j})$ is the probability that branch value $\hat{\omega}_{i_n}$ is decoded given that the codeword for branch value $\hat{\omega}_{i_j}$ was transmitted. The summation is performed over all possible received codewords. This conditional probability is a function of the expected BER of the channel and is easily derived using the binomial distribution. It is worthy to note that (5) reduces to (4) for the noiseless channel. By expanding (5) it is easily shown that a considerable amount of precomputation is possible. In fact, by using lookup tables the computational complexity of the joint source and channel cost function (5) exceeds that of the source coder cost function (4) by a single addition [12].

The chosen system codes LSPs at 33 b/frame while exhibiting remarkable immunity to a wide range of channel BERs. In Fig. 1 the system is compared to the 34 b/frame Federal Standard 1016 scalar LSP quantizer [16] using the rms log spectral distortion measure. The performance of the joint source and channel trellis LSP coder is seen to exhibit high immunity to a wide range of BERs and is not characterised by a rapid rise in distortion at very high BERs.

By combining source and channel coding in this way, the problem of apportioning bits to the source and channel code is shifted to one of deciding which channel BER to design for. If designed for high BERs, performance is reduced when operating with noiseless channel conditions and conversely if designed for low BERs the code deteriorates rapidly with high rates of channel error. As a compromise, the designed expected channel BER was set to 0.04 which offers good performance at low BERs yet performs without severe degradation at high BERs (greater than 0.01).

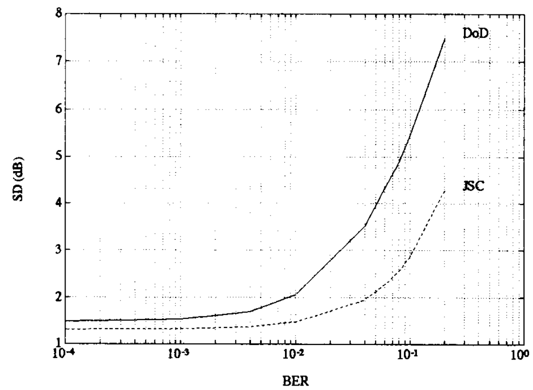

Figure 1. Performance of Joint Source and Channel Trellis LSP coder compared to Federal Standard 1016 Scalar LSP Quantizer.

4.0 RESULTS

The encoded MBE model parameters are fundamental frequency (pitch), joint source and channel coded LSP information, LPC spectrum gain, and the voiced/unvoiced decisions. The fundamental frequency and LPC spectrum gain are uniformly and log quantised respectively using 8 bits each. The voiced/unvoiced decisions are binary values and require no further quantisation.

The LPC spectral information is protected using joint source and channel coding techniques and requires no Forward Error Correction (FEC). The fundamental frequency and LPC spectrum gain are partially protected using a (23,12) Golay code. Table 1 presents the bit allocation of the coder.

Parameter	Bits/Frame	Bits/s
LPCs	33	1100
V/UV decisions	12	400
Fundamental Freq.	8	266.66
LPC Gain	8	266.66
FEC	11	366.66
Total	72	2400

Table 1. Coder Bit allocation

The coder was tested in random error channels with bit error rates of 0, 0.005, 0.01, 0.02, 0.05 and the results

evaluated using paired comparison tests (Table 2) on a 3 second male utterance.

Condition	Mean
Original utterance	4.83
Coded, BER = 0	3.33
Coded, BER = 0.005	3.17
Coded, BER = 0.01	1.83
Coded, BER = 0.02	1.83
Coded, BER = 0.05	0

Table 2: Paired Comparison Test Results

The coder shows little perceptual degradation for bit error rates up to 0.01. At bit error rates of 0.01 and 0.02 the coder quality is still acceptable, however the Golay code starts to break down at bit error rates greater than 0.05, significantly affecting the speech quality.

5.0 CONCLUSIONS

A robust 2400 bits/s MBE coder is presented. The coder employs a non-redundant spectral envelope encoding technique based on joint source and channel trellis coding of LSPs. This technique combines high source compression while minimizing the distortion due to the channel. The coder is robust to a wide range of BERs and is not accompanied by a rapid breakdown at very high BERs. Remaining critical parameters, not amenable to trellis coding, are protected by traditional FEC methods. The coder performs remarkably well over a wide range of BERs, with intelligibility maintained to BERs up to 5%.

6.0 REFERENCES

[1] Digital Voice Systems, "Inmarsat-M Voice Codec, Version 2", Inmarsat-M specification, Inmarsat, London, Feb. 1991.

[2] D. W. Griffin, J. S. Lim, "Multiband Excitation Vocoder", IEEE Trans. ASSP, vol ASSP-36, pp. 1223-1235, Aug. 1988.

[3] J. C. Hardwick, "A 4.8 Kbps Multi-Band Excitation Speech Coder", S.M. Thesis, E.E.C.S. Department, M.I.T., May 1988.

[4] M. S. Brandstein, "A 1.5 Kbps Multi-Band Excitation Speech Coder", S.M. Thesis, E.E.C.S. Department, M.I.T., 1990.

[5] D. Rowe, W. Cowley, A. Perkis, "A Multiband Excitation Linear Predictive Hybrid Speech Coder", Proc. Eurospeech '91, pp. 239-242, Genova, Italy, Sept. 1991.

[6] S. Yeldener, A. M. Kondoz, B. G. Evans, "High Quality Multiband LPC Coding of Speech at 2.4kbit/s", Electronic Letters, vol. 27, No. 14, July, 1991.

[7] J. Makhoul, "Linear Prediction: A Tutorial Review", Proc. of the IEEE, vol. 63, pp. 561-580, April 1975.

[8] D. O'Shaughnessy, "Speech Communication, Human and Machine", Chapter 8, Addison-Wesley, 1987.

[9] N. Farvardin, "A study of vector quantization for noisy channels," IEEE Trans. Inform. Theory, vol. IT-36, pp. 799-809, Jul. 1990.

[10] E. Ayanoglu and R.M. Gray, "The design of joint source and channel trellis waveform coders," IEEE Trans. Inform. Theory, vol. IT-33, pp. 855-865, Nov. 1987.

[11] P. Secker, A. Perkis, "Joint Source and Channel Coding of Line Spectrum Pairs', Proc. Eurospeech '91, pp. 909-912, Genova, Italy, Sept. 1991.

[12] P. Secker, A. Perkis, "Joint Source and Channel Coding of Line Spectrum Pair Parameters', Speech Communication, submitted for publication.

[13] P. Itakura, "Line Spectrum Pair representation of linear prediction coefficients of speech signals", J. Acoust. Soc. Am., vol. 57, 1975.

[14] F. Soong and B. H. Juang, "Line spectrum pair (LSP) and speech data compression," Proc. ICASSP-84, pp. 1.10.1-1.10.4, 1984.

[15] J.G. Dunham and R.M. Gray, "Joint source and channel trellis encoding," IEEE Trans. Inform. Theory, vol. IT-27, pp. 516-519, Jul. 1981.

[16] R. Fenichel, "Proposed Federal Standard 1016 - second draft," National Communications System, Office of Technology and Standards, Washington, DC 20305-2010, Nov. 13, 1989.